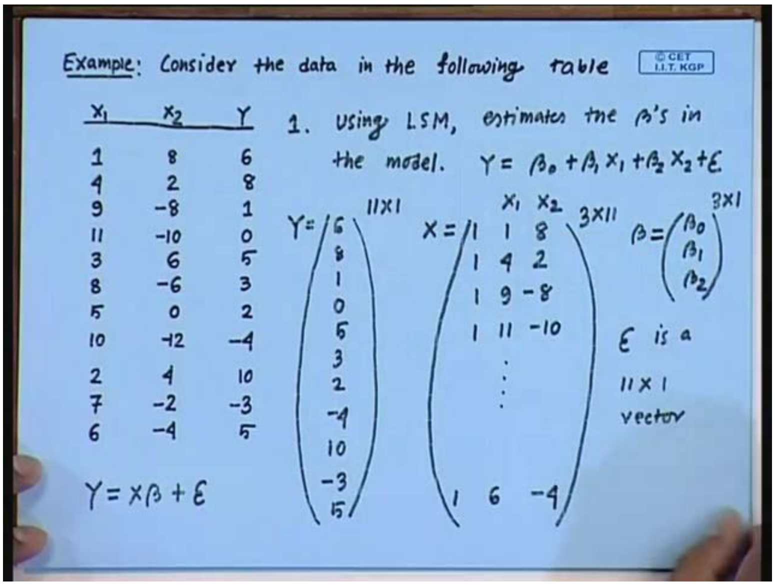
Regression Analysis
Prof. Soumen Maity
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture - 8
Multiple Linear Regression (Contd.)

Hi, this is my third lecture on Multiple Linear Regression. Today we will consider one example and this example will enable as to illustrate the theoretical concepts, we be mentioned in the last 2 classes. Well in last 2 classes, we have learned you know how to feet multiple linear regression model and when the model has benefitted has on being constructed, next job is to you know the test the statistical significance of the fitted model. And we know how to test a, the significance of the fitted model using the global test and also using the partial test.

(Refer Slide Time: 02:02)



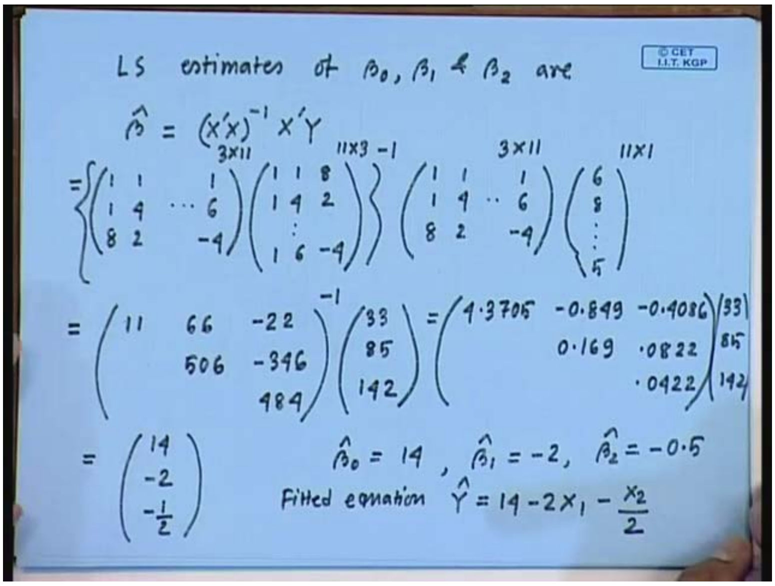
So, we considered the data in the following table, here you have a 2 regressive variable X 1 and X 2, and 1 response variable Y and here the observation corresponds to X 1, X 2 and the response variable Y. So, this one is the basically you know this is the first observation on a regressor 1, this the first observation on regressor 2, and this is the first observation on the response variable Y.

Well so what is the first requirements is that using least squares method estimates, the parameters meters in the model, so the model here is you know, Y is equal to beta naught plus beta 1 X 1 plus beta 2 X 2 plus get the epsilon. So, if you estimate that the unknown parameters beta naught beta 1 and beta 2 well, show first we construct the following matrix is matrix Y, which is you know basically the matrices of observations, so this is 6 8 1 0 5 3 2 minus 4 10 minus 3 5.

So, this is the vector of a observation and this have 11 cross 1 vector, because we have 11 of observations here, now we will construct the matrix X, this X is equal to 1 1 8 1 4 2. So, this column is corresponds to the regressor X 1, this one is corresponds to the regressor X 2 and this where 1 9 minus 8, 1 11 minus 10, this way we go up to 1 6 minus 4. So, this one is 3 cross 11 matrixes and the next one is the beta vector, which is the basically the vector of parameters, this consists to beta naught beta 1 beta 2, so it is 3 cross 1 vector.

And also, we defined we know epsilon, epsilon is a 11 cross 1 vector, which one is you know, this is the vector of errors, show the model, now can be returned in the matrixes form like Y is equal to X beta plus epsilon well and we need to estimate the unknown parameters, using least squares method.

(Refer Slide Time: 07:15)



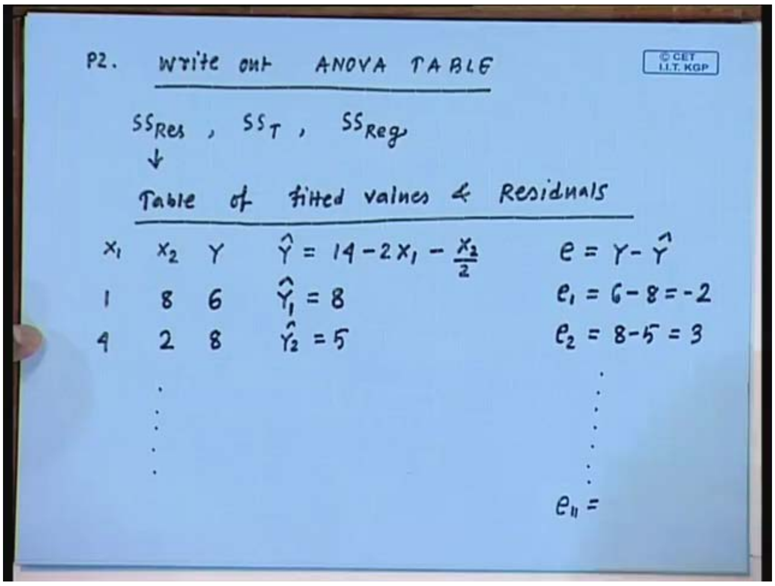
Well so, L S estimates of beta naught beta 1 beta 2 and beta 2 are beta hat, which is equal to X prime X inverse X 1 Y, so this one is you know, 1 1 8 1 4 2 1 6 minus 4, this one my X prime and X is 1 1 8 1 4 2 1 6 minus 4. So, X prime X inverse X prime Y, so 1 1 8 1 4 2 1 6 minus 4, this is X prime and Y is 6 8 5 right. So, see you know, this one is X is 3 cross 11 matrixes X prime is 11 cross 3 matrixes and this one is X prime and so this is 3 cross 3 by 11 matrixes and this one is 11 cross 1 matrix well.

So, this one is you know, you can check that X prime X is equal to 11 66 minus 22 506 minus 346 484 and it semantic matrixes and it is a symmetric matrixes. So, this is my X prime X, so X prime X inverse and X prime Y is 33 85 142, so you can check that will now the universe of this matrix's is equal to 4.3705 minus 0.849 minus 0.4086, well so this inverse is also of course, it semantic matrix's, so this is the inverse of X into X prime Y, which is equal to 33 85 142.

So, what we got is that, you can check that, this is equal to 14 minus 2 minus half, so beta are not hat is equal to 14 beta 1 hat is equal to minus 2 and beta 2 hat equal to minus 0.5. So, our fitted equation is fitted equation is Y hat equal to 14 minus 2 X 1 minus X 2 by 2, well so once, we have the fitted equation and you known next, we need to check how useful is fitted equation, that means, we need to test the statistical significance of the fitted equation.

Basically, first we go for the global test, we which test the hypothesis that, beta 1 is called to beta 2 is equal to 0, that means, there is no linear relationships between the responsive variable and regressor variable. And will you use the you know and you are approach to test that statistical significance of the fitted equation, well so far that, we you know first we need to write the down that ANOVA table, let me to that first.

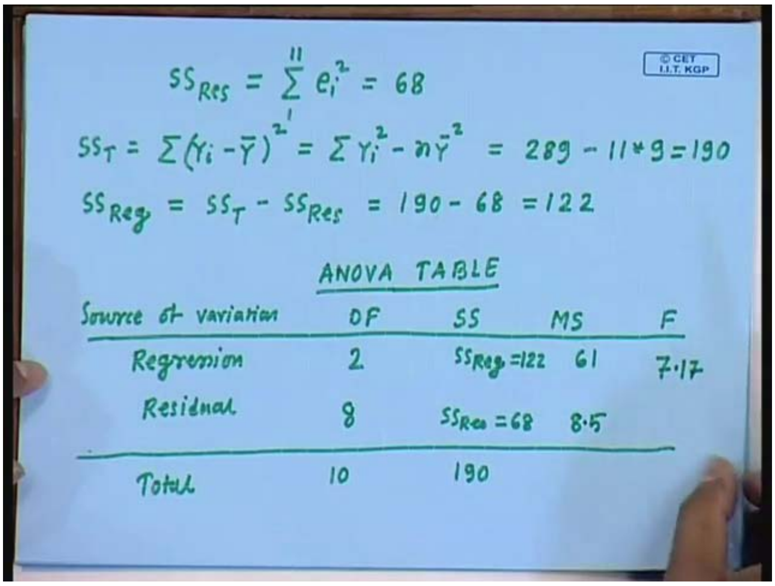
(Refer Slide Time: 14:18)



Well so, the next problem you know p 2, I will say the second problem, you write out ANOVA table, well so, what we to a compute here is that, you need to compute S S residual S S total and then S S regression can be obtained from these 2, I mean S S T minus S S residual is equal to S S regression. Now, you are not at the compute, you know S S residual, first to in need to compute the residuals well, the table for a we make a table of fitted values and residuals, we know what these X 1 X 2 Y, you know, what is Y hat, Y hat is equal to 14 minus twice X 1 minus X 2 by 2 right.

So, once we have Y and Y hat, we can compute e, e is equal to the residual is equal to Y minus Y hat, let me just explaining 1 2 a observation like first observation is 1 8 and the response value is 6, so the corresponds to 1 and the 8, the fitted value of responsive variable is equal to 8 you can check that. So, the first residual, you know basically, this is Y 1 hat, so e 1 is equal to Y 1 minus Y 1 hat, Y 1 is equal to 6 and Y 1 hat is equal to 8, so the first residual is minus 2, similarly 4 2 8, this is second observation. So, Y 2 hat is equal to 5 and you will e 2 is equal to 8 minus 5 is equal to 3. So, this you know compute all the residuals, so this way, you can go up to e 11, because you have 11 observations and once you have all the residuals, you know, you can compute the S S residual now.

(Refer Slide Time: 17:49)



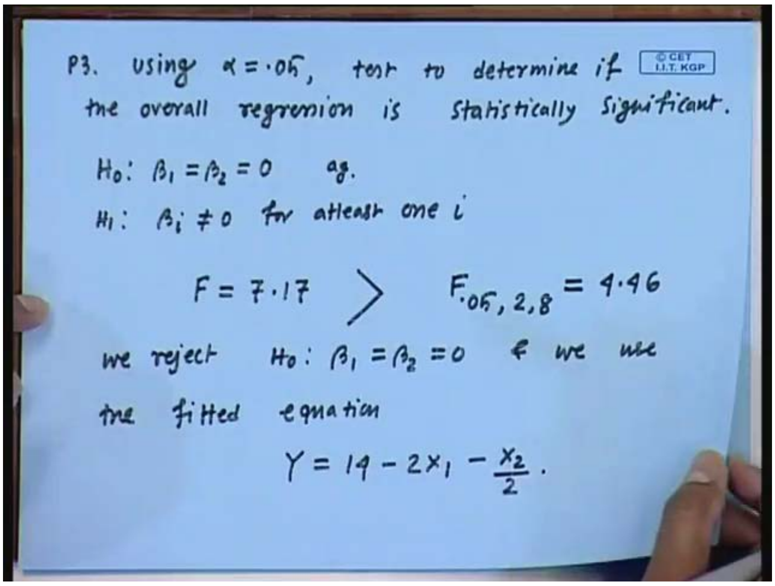
So, S S residual is recalled to summation e i square, i is from 1 2 right and the value of this one, you can check that, this is 68, next we go for S S total, S S total is equal to summation Y i minus Y bar whole square, which is nothing but summation Y i square minus n Y bar square right. This value is equal to 289 minus n is 11 and Y bar is 3, so 3 square is equal to 9 and this is total is equal to 190 and once I have the S S total and S S residual, I can compute S S regression, which is equal to S S total minus S S residual 190 minus 68, which is equal to 122 right.

Now, we can have the ANOVA table, source of variation non degree of S S M S and the F follow will the source of variation explained by the regression model, residual and the total variation well. The degree of freedom of total S S total is 10, because you know we have 11 observations and 1 degree of freedom, we lose, because the constant that summation Y i minus Y bar is equal to 0. So, S S total as degree of freedom 10 than S S regression as degree of freedom 2 and then S S residual degree of freedom 8.

Let me explain here, tou know S S regression as degree of freedom, which is 8, which is equal to 11 minus 3, we are losing the 3 degree of freedom, because of the 3 constrain on residual, because the residual e i, it satisfies you know, 3 constrain. The fast constrain is summation e i is equal to 0, the second constrain is summation e i X 1 is equal to 0, X 1 is the first regressor and the 3rd constrain is summation e i X i 2, for the see here. Well so, the residual as degree of freedom 8, on the regression has the freedom 2, we call it S S residual, this is called S S regression.

So, the S S regression value is equal to 122, S S residual is equal to 68 and this one is 190 and the M S value is 61, so 122 by 2 and MS residual is 68 by 8, which is equal to 8.5 and the if statistic value is equal to 7.17. So, this is the ANOVA table, the next we move to the next problem, which is problem 3.

(Refer Slide Time: 23:46)

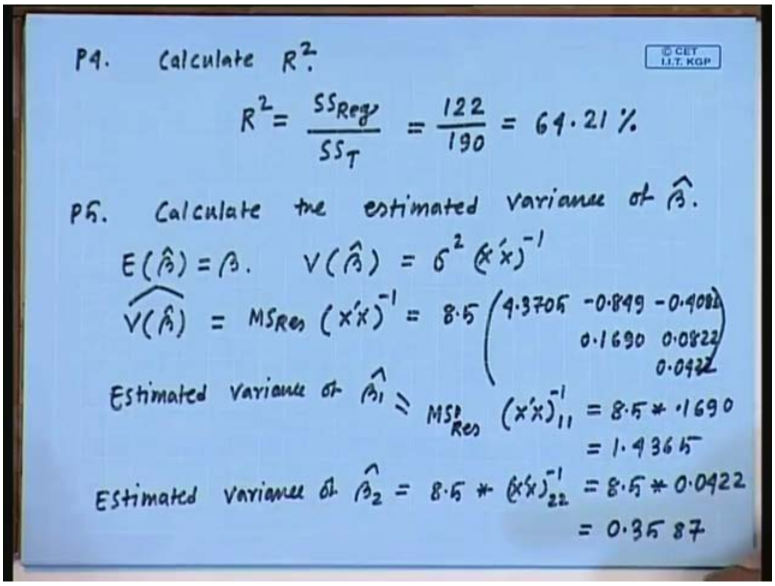


I mean requirement 3 you can say, using you know, alpha is equal to 0.05, that means, the level of significance is 0.05, you test to determine, if the overall regression is that statistically. So, the meaning of this one is that, whether the overall regression is statistically significant, that means, we need to go for the global test, that means, we need to test the hypothesis that beta 1 is equal to beta 2 is equal to 0, this is the null hypothesis against the alternative hypothesis that, H naught is not true.

So, the basically, we need to test the hypothesis H naught beta 1 is equal to beta 2 is equal to 0, this says that there is no linear relationship between Y and the regressor variable against the alternative hypothesis. H 1 is says that, beta i naught equal to 0, for at least 1 i, to test this hypothesis, we what we do is that we compare, you have the F statistic values. So, we compared the F value, observed value, which is equal to 7.17, we compare this observed value with the F tabulated value, F 0.05, the degree of freedom is 2 8, from the statistical table, you can check that these value is equal to 4.46.

So, what we observed that, you know, the observed F value is greater than the tabulated value, so we reject the null hypothesis, H naught which says that beta 1 is equal to beta 2 is equal 2 0. And we use that means, the fitted equation is significant use the fitted equation Y equal to 14 minus 2 X 1 minus X 2 by 2. Well, so what the results of this test is the global test, it says that, fitted equation is statistically significant, that means, there is you know linear relationship between Y and any at least one of the response one of the regression variable.

(Refer Slide Time: 28:00)



Well so, next requirement are problem 4th and should I say are you calculate R square, that is the you know, coefficient of determination, this is known basically R square is equal to S S regression by S S total. Well our S S regression value is equal to 122 and S S total is 190, so this is this R square is 1 parameter, which major, we know sort of the performance of the fitted equation, it majors the proportion of variability in, I about Y bar that is explained by the fitted equation.

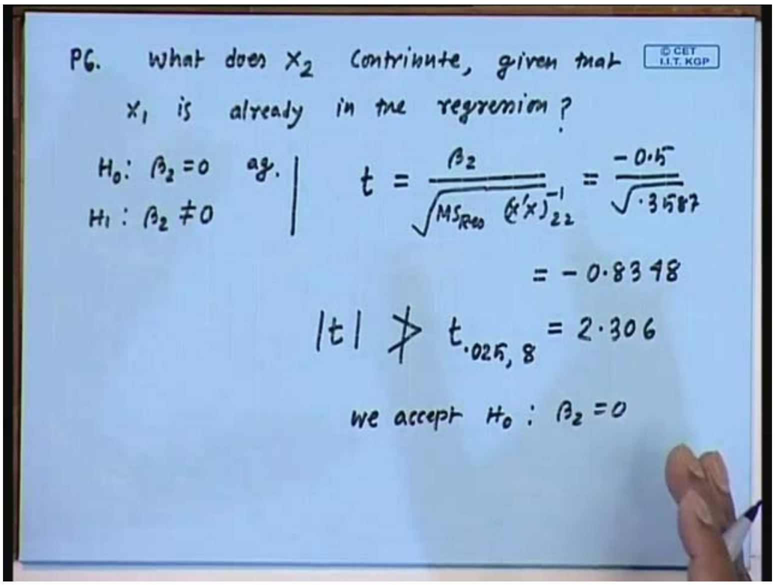
Well so, the proportion here is 122 by 190, which is equal to the 64.21 percent, which is not that good you know, that means, the 64 percent of the total variability in the in the response variable has been explained by the 2 regressor variable. So, it is not that, we know not that good, well so next, we move to the next problem 5, it says that, you know calculate the estimated variance of beta hat, what we know is that, you know beta hat is an unbiased estimator of beta, you know that expected value of beta hat is equal to beta.

And also, we know the that the variance of beta hat is equal to sigma square X prime X inverse well. And what you want is that, do you want estimated value of this variance well. So, here to estimates the variance of beta hat, already you know, sigma square is not known, so known, so and also you know that M S, M S residual is an unbiased estimator of sigma square, the basically will be replacing sigma square by M S residual well all.

So here, I can write you know variance of beta hat are the estimate of this one is equal to M S residual into X prime X inverse, which is nothing but 8.5 is the M S residual, 8.5 into this matrixes X prime X inverse, which is equal to 4.3705 minus 0.849 minus 0.4086 right well. Now, estimated variance of beta 1, beta 1 hat is equal to M S residual X prime X inverse 1 1 is the element, we know, I have used the notation for the variance of beta j hat j j here.

So, the meaning of this one is I call this one is 0 0 element, because this one is corresponds to beta naught, this element corresponds to this is the variance co variance matrixes for beta naught hat beta 1 hat and beta 2 hat, the variance of beta 1 hat is M S residual into the 1 1 at the element of this matrix. So, this one is nothing but 8.5 into 0.1690, which is equal to 1.4365, similarly and the estimated variance of beta 2 hat is equal to 8.5 into X prime X inverse, 2 2 at a element, which is nothing but 8.5 in to 0.0422, which is equal to 0.3587.

(Refer Slide Time: 35:21)



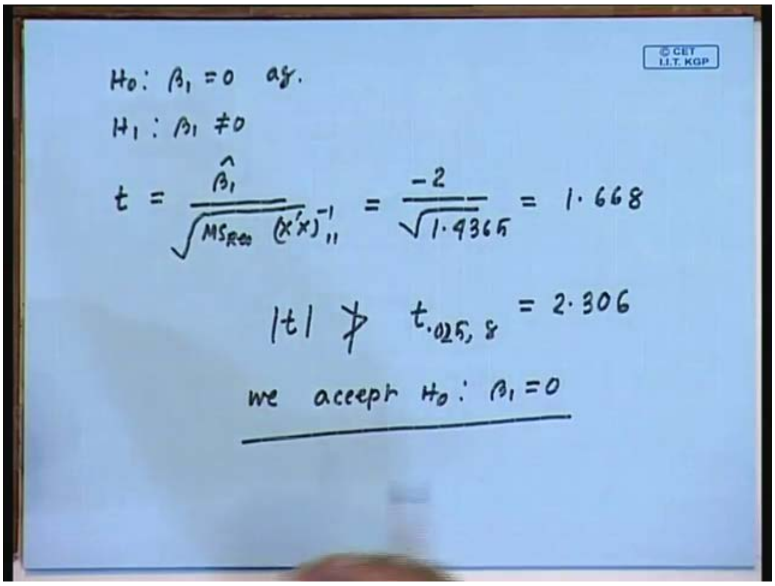
Well so, next problem is P 6, it says that, what does X 2 contributes given that X 1 is already in the regression model, well so, it basically says you know, the contribution of the second regressor that is X 2, in the presence of the first regressor in the model, you know, if you recall partial test partial. So, basically here you have to test that, test the null hypothesis that, which whether beta 2 equal to 0 against the alternative hypothesis disease that, beta 2 naught equal to 0.

Well sure the meaning of this hypothesis is that, what is the contribution of the second regressor X 2 in the presence of the first regressor X 1 in the model. So, the hypothesis to test this is beta 2 is equal to 0, against the alternative hypothesis H 1, which says that beta 2 is naught equal to 0. And we know that, test statistic to test this hypothesis is t, which is equal to beta 2 by M S residual X prime X inverse 2 2th element and you know the meaning of this one.

So, this is basically variance of beta 2 estimate of the variance of beta 2 and this quantity is equal to minus 0.5, just now we calculated this is 0.3587 and this is going to be minus 0.8348. Now, here t follow, t distribution the degree of freedom 8, so you find out the tabulated value of t point is 0 2 5 that is 2 sided test, if the degree of freedom 8. This value is equal 2.306, which is show the observed value S, the modern the observed value is not greater than the tabulated value, show the conclusion is that, we accept H naught, it says that beta 2 is equal to 0.

So, the meaning of we know, what we concluded is that X 2 does not have a any contribution in the regression model, in the presence of X 1. So, let me tell the other thing, we also ward as X 1 contributes in the presence of X 2 in the model, so for that, I need to test.

(Refer Slide Time: 40:01)



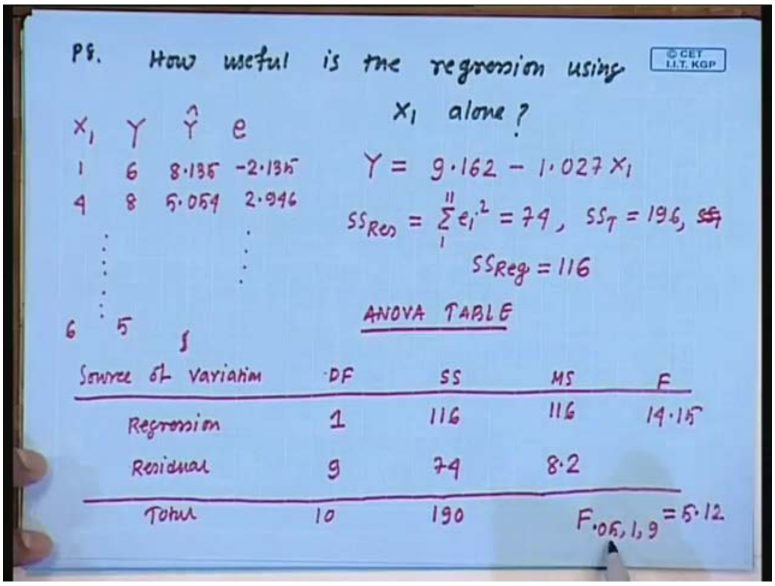
Similarly, I need to test H naught, we says that beta 1 naught is equal to 0, against H 1 beta 1 is naught is equal to 0 and the test statistic for this one is the t, we is called to beta 1 hat by M S residual X prime X inverse 1 1 at the element. This is equal to minus 2 by 1.4365, which is going to 1.668 and again you know this t is naught greater than R equal to t tabulated value, t 0.00258, which is equal to 2.306.

So, the conclusion here is that, we are again we accept H naught that beta 1 equal to 0, that means, you know the conclusion of this partial test says that, there is that X 1 is not significant X 1 does not significant contribution to the model in the presence of X 2. And also, the first partial test that is beta 2 is equal to 0, that also is accepted, so the conclusion of other partial test says that, X 2 is also not significant in the presence of X 1 in the model.

So, this is you know whereas, the global test says that, X 1 and X 2 is significant, that means, the model is significant, but individually X 1 is not significant in the presence of X 2 and similarly X 2 is not significant in the presence of X 1. So, this is one nice example know, global test is reject, you know global test says that, X 1 X 2 significant to explain the variability in Y, why that whereas, the partial where other hand partial test says that, either X 1 is significant in the presence of X 2 nor X 2 is significant in the presence of X 1.

So, the here you know this example, explain the problem of multi co linearity, in the regressor variable, so anyway, I will be going to talk about multi co linearity later on. So, this is the results of partial test, so what have to do the next is that, I want to test how useful.

(Refer Slide Time: 43:27)



Problem 8, how useful is the regression using X 1 alone, that means, if you only consider X 1, in the model then how much of the variability Y is explained by X 1 that is what we want to check. So, what I am doing is that, I am just removing the second variable X 2, I will be only talking about, X 1 and Y, that means, I have the observation like 1 6 4 8 6 5 and I want to fetid simple linier regression between Y and X 1, you can check that, you know out to linear regression model.

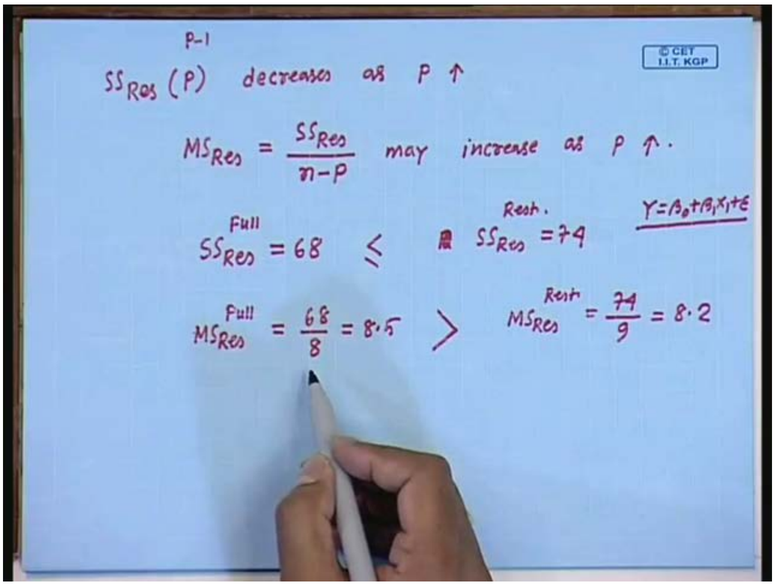
So, it can check that y is equal to fetid model Y is equal to 9.162 minus 1.027 X 1, so once you have that fetid model, you can compute, you know Y hat you can compute the residual for example, Y 1 hat is equal to 8.135 and then e 1 is equal to minus 2.135. Similarly, the for the second observation, it is 5.054 and the e 2 is 2.946, so you do eat for all the observations and you compute S S residual, which is equal to summation e i square 1 to 11, which is equal to 7 4. And S S T is same S S T is equal to 190 and S S, S S regression sees then equal to 116.

Now ANOVA table source of variation, degree of freedom S S M S and the of statistics to the regression residual, so the degree of freedom here, were total degree of freedom is strain, regression as degree of freedom 1, the residual as degree of freedom 9, I will hope you understood, why it is 9, S S regression is 116 residual is 74 total is 190. So, M S 116 and M S residual is equal to 8.2 and the F value is 14.15, now you check you know here, this F follows F distribution with degree of freedom 1 and 9.

So, you find out the value of F is 0 5 1 9 from the table, it is equal to 5.12, so the observed value of F is greater than the tabulated value, which is equal to 5.12, that means, you know the regression the fetid model is the significant. And also one more thing to I observed here, you know here, before this S S regression was 122, and now, that S S regression, I mean the S S regression involving 2 regressor was 122 and S S regression involving only 1 regressor is 116.

So, perhaps you know X 1 is more capable to explain the variability in Y compared to X 2, because know of 122, which was the S S regression before involving 2 regressor X's 1 and X2. Now you keep only one regressor in the model 6 1 then it is explaining, we know almost is the same well, now one more thing, discus here is that.

(Refer Slide Time: 49:59)



Let me defined one thing, S S residual P, so this the meaning of this one is that the residuals some of square and in they are P minus 1 regressor in the model and of course, i e you know this S S residual involving P minus 1 regressor in the model, this decreases as P increases. So, if you involved more regressor in the model, then S S residual decreases, well but this is not true, for M S residual, M S residual is equal to S S residual by n minus P, this may increase as P increases.

You know, why it is so, just you considered the previous examples, there you know S S residual, initially it was 68, when we had both the regressor X 1 and X 2 in the model. So, I will say this S S residual for the full model and this one is less than the S S residual for the restricted model, I means the model Y equal to beta naught plus beta 1 X 1 plus epsilon. So, for this model also, you have fitted the model, we check that the S S residual for this one is equal 74 right.

So, this one has more regressor that is why the S S residual is less compared to this model this model as already one regressor, this model as the full model and 2 regressor's. Now, you compute the M S residual for the full model, that means, 2 regressor's is equal to 68 by degree the freedom of a 8, which is equal to 8.5, for this restricted model, you compute M S residual for the restricted model, which is equal to 74 by 9, which is equal to 8.2.

So, the m s residual for the full model is greater than the M S residual, for the restricted model, so this explains you know, I said that S S residual always decreases has been increases, for the same is not true, for M S residual. The reason is here, you know the increase in M S residual, occurs when the reduction in you know, if you increase one more variable in the model of course, S S residual decreases.

But, the reduction in S S residual here, for I mean for adding one more regressor in the model is not sufficient to compensate the loss of one degree of freedom in the denominator show here. If you add one more regressor the model this one is model involving one regressor X 1 only, this one is the model involving 2 regressor's X 1 and X 2.

But, the reduction in S S residual is not enough to compensate, you know the loss of one degree of freedom in denominator that is why, the M S residual for the full model, I mean the model involving more regressor is more than the M S residual in the restricted model. So, this is one thing here going to use later on like a M S residual, as a parameters to selector the best model, you will learn those things later on.

Thank you and let us see.